AD_____

Award Number: W81XWH-11-1-0487

TITLE: Development of Advanced Technologies for Complete Genomic and Proteomic Characterization of Quantized Human Tumor Cells

PRINCIPAL INVESTIGATOR: Dr. Robert Moritz

CONTRACTING ORGANIZATION: Institute for Systems Biology
Seattle, WA 98103

REPORT DATE: July 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2014	Annual	15 Jun 2013- 14 Jun 2014

4. TITLE AND SUBTITLE
Development of Advanced Technologies for Complete Genomic and Proteomic Characterization of Quantized Human Tumors

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0487

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Robert Moritz

E-Mail: Robert.Moritz@systemsbiology.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Systems Biology

Seattle, WA 98109

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
With the establishment of Glioblastoma (GBM) cell lines from GBM patient's tumor samples and quantized cell populations of each of the parental GBM cell lines, we have completed most of our major aims of this project. We will continue in our efforts in the development and analyses of these quantized cells and develop the genomic and proteomic technology to interpret specific transcriptome and proteome signatures. Whole genome sequencing from two families of GBM patients are now well established and from the basis of the molecular characterization of the tumor development and signatures presented by these tumors for the development of diagnostic signature panels. We have established efficient secretome cell culture conditions to enable proteomic analysis of these quantized cell sub populations and have begun to assemble the protein signatures of GBM tumors underpinned by the comprehensive molecular characterization of these tumors through family WGS, transcriptomics of single cells and secreted protein analysis. This program will ultimately provide new technical capabilities through broad based molecular characterization at the genomic and proteomic analysis of GBM patients and their immediate families and has already provide significant results. This final effort in the molecular characterization of a second patient and their family and the establishment and deployment of a proteomic signature for GBM diagnostics will be completed within the next 12 months.

15. SUBJECT TERMS
Human cohorts, Glioblastoma, Genomic, Proteomic, Single-cell technologies, Hypothesis-driven, integrative systems approach, Early diagnosis,Patient stratification,Blood protein biomarkers,Quantized cell populations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	22	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

We have developed quantitative tools for direct clinical application to human cohorts with glioblastoma classified cancer. This program promises to deliver important insights to cancer mechanisms (disease-perturbed networks), as well as blood biomarkers to assess progression and stratification of human glioblastoma. This proposal will significantly advance genomic, proteomic and single-cell technologies, enabling the commencement of hypothesis-driven integrative systems approaches to disease (cancer). To this end, we have developed new strategies for advanced genome sequencing, new technologies for the analyses of transcriptomes, miRNAomes and single cells as well as multiplexed quantitative protein measurements including the measurement of isoforms, and post-translational modifications. The tools proposed here will be generally applicable to all cancer-based studies, as the nature of the tool development is designed to identify and quantify DNA, RNAs, proteins and cells, challenges ubiquitous to all human disease systems.

To complete these tasks, we have used a logical approach developed with the following aims:
Specific Aim 1. Isolate up to 1000 cells from each of five human glioblastomas and quantify initially 500 different transcripts from each cell (transcription factors, CD molecules, relevant signal transduction pathways, etc). Determine whether computational analyses can classify these cells into discrete quantized cell types.
Specific Aim 2. Sort the disassociated tumor cells from several glioblastomas into their quantized cell populations using cell sorting/CD antibodies to each quantized cell type for functional analyses and establish primary cell lines. These cells are characterized morphologically and are used for the suite of molecular analyses—at the genome, transcriptome, miRNAome and selected proteome levels.
Specific Aim 3. Assess 20-40 candidate blood biomarkers in the bloods of 100 glioblastoma patients with regard to their ability to stratify disease, assess disease progression and predict at an early stage the reoccurrence of the glioblastoma (early detection). Eventually we will use these biomarkers to assess the effectiveness of therapy.
Specific Aim 4. Ten to 20 cells from each major quantized glioblastoma cell type from two patients have been used to determine the complete genome sequences. We have also determinde the normal genome sequences of each patient and their family members to enable the Mendelian-based error correction process recently described in our recently published Science paper (*1*). The mutations defined in GBM tumors are analyzed against quantitative changes in the transcriptomes, miRNAomes and proteomes and against the relevant biological networks.
Specific Aim 5. Analyze the quantized cell populations for their responses (transcriptome, miRNAome, etc) to the perturbations of key glioblastoma-relevant molecules (e.g. nodal points in networks) by RNAi perturbations as well as their responses to glioblastoma-relevant drugs

and natural ligands. These assays will be carried out in the laboratory of our collaborator Dr. Charles Cobbs and Dr. Parvinder Hothi at Swedish Neuroscience Institute.

The expected outcomes and deliverables of this innovative program are: **1**) a deeper understanding of human glioblastoma disease mechanisms; **2**) the establishment of new blood protein biomarkers for use in early diagnosis, stratification of glioblastoma tumors, assessment of the progression of a glioblastoma tumor, identifying biomarkers for assessment of effectiveness of drug treatment and detection of reoccurrence at an early stage; **3**) new strategies for genomic sequencing of quantized cancer cells and their normal counterparts to identify cancer-driver mutations; **4**) new technologies for transcriptome, miRNAome, proteome and single-cell analyses, and **5**) the creation of quantized glioblastoma cell lines that can be used for general molecular characterization as well as to assess the biology of this cancer (drugs, RNAi's, natural ligands) and the effectiveness of existing drugs in reacting with these cell types.

BODY

Aim 1, 2 & 4. Our work is linked with the Ivy Center for Advanced Brain Tumor Treatment at the Swedish Neuroscience Institute (SNI) collaborative group (CA100459P1, Award Number W81XWH-11-1-0488, Swedish Health Services, Dr. Charles Cobbs) to provide cells from human glioblastoma tumors (GBM) excised during surgery at the Swedish Neuroscience Institute (SNI) by our original collaborator, Dr. Gregory Foltz (deceased). During the past year, we continued to collect GBM tumor samples to establish the patient cohort available for molecular analysis, genome sequencing, and quantitative assays. To date, the SNI has collected tumor tissue eligible for this program from over forty GBM patients. We established several primary GBM cell lines from patients undergoing tumor resection at SNI with two patients providing complete family consent for WGS of available members. We confirmed in these cultures stem cell phenotype by functional assays of self-renewal, differentiation potential, and tumor propagation *in vivo* where average tumor volume (n = 5) in immuno-compromised mice six weeks after implantation of GBM-patient derived cells increase by average 5-fold. The final two patients selected for complete molecular analysis of GBM-derived primary and quantized cell culture are designated SN291 and SN 243. These patients have the requisite consenting family members to complete the available samples needed to satisfy our aims (Specific Aims 1 &2) and have been used for molecular analyses at the genome, transcriptome, miRNAome and proteome levels. For each of these established cell lines, a number of single cell clones have been successfully established from the corresponding parental cell lines.

SN#	Gender	Age	Histopathology	Resection	Subtype	MGMT	Chemotherapy	Radiation	Survival (days)	Status
143	Male	75	GBM (Gliosarcoma), grade IV	Left Temporal	Mesenchymal	Unmethylated	Not available	Not available	323	Deceased due to tumor progressi
186	Male	76	GBM, grade IV	Right Temporal	Proneuronal	Unmethylated	140mg TMZ, over 11 weeks (concurrent with radiation).	IMRT, 4500 cGy in 25 fractions, over 6 weeks.	459	Deceased due to tumor progressi
243	Male	57	GBM, grade IV	Right Frontal	Proliferative	Methylated	160mg TMZ, concurrent, 6 weeks; 400mg TMZ, maintenance 5x/mo, 38 weeks; 160mg TMZ, maintenance 21x/mo, 8 weeks; 400mg TMZ, maintenance 5x/mo, 8 weeks.	IMRT, 4140 cGy in 23 fractions, concurrent, 3 weeks. IMRT, 1800 cGy in 10 fractions, boost, 3 weeks.	N/A	Alive, 3 yrs 2 months post surge
291	Female	63	GBM, grade IV	Right Parietal	Mesenchymal	Methylated	150mg TMZ, every 2 weeks 5 days cycle, for 54 weeks.	IMRT, over 8 weeks. Stereotactic, 2500 cGy in 5 factions, 1 week.	N/A	Alive, 2 yrs 2 months post surge
348	Female	49	GBM, grade IV	Right Frontal	Not determined	Unmethylated	105mg TMZ, concurrent, 7 weeks.	IMRT, 5940 cGy in 33 fractions, 7 weeks.	123	Deceased due to tumor progressi

Table 1. Clinical diagnosis, treatment and survivability of GBM patients which have provided samples to this program.

We have established quantized cell populations from the primary glioblastoma tumors of SN291 and SN243 for whole genome sequencing (WGS), transcriptomics, and proteomics analyses of these quantized cell populations for novel biomarker discovery. To complete this aim, we need blood and PBMC cells from both patients and their family members. Our clinical collaborators at SNI completed recruitment for both patients and their consenting families with all specimens required for full molecular analysis over this last year. For quantized cell

populations, we developed single cell clonal culture techniques and integrated single cell sorting using the BD FACS Aria II into precoated 384-well plates. Approximately 60% of these sorted cells formed colonies (>100 cells) and were collected and frozen ready for further analysis. For each primary tumor line, we established dozens of clonal cultures which exhibited distinct morphological phenotypes, and each clone presumably carries a uniform genome, thus is ideal for subsequent WGS analysis. For example, we established 12 new single cell clonal cultures from patient SN291, using the same protocol that we have established for earlier GBM patient samples and described in the previous quarterly and annual reports. We finalized our selection to 5 clones from each patient which had differing phenotypes and selected these for subsequent 'omic analysis.

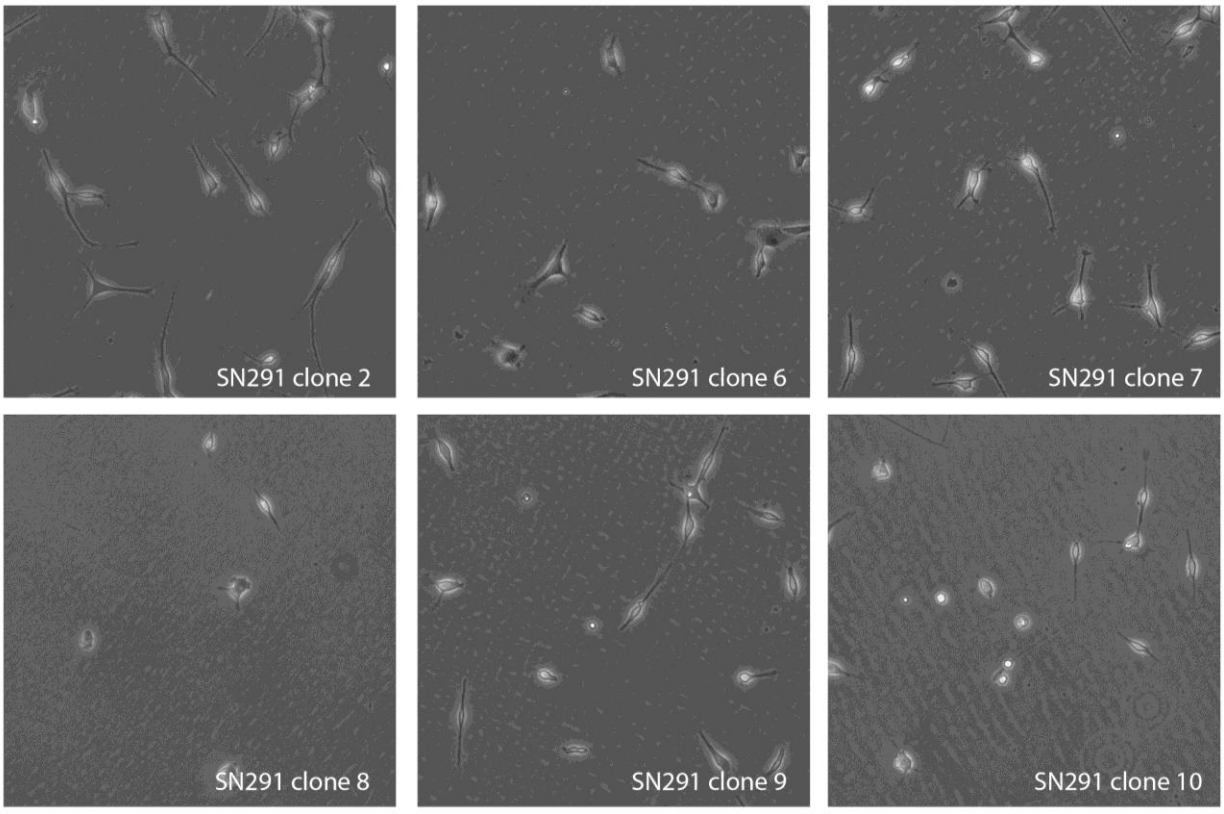


Figure 1. Establishment of new single cell clonal cultures from GBM patient SN291. A total of 12 clonal cultures have been generated from this patient sample. Culture protocol used from Pollard SM et al., 2009, Cell; Stem Cell. Briefly, these cells were cultured on plates coated with laminin and grown under Serum-free conditions with stem cell media and addition of B27 and N2 supplement, growth factors: EGF and bFGF. The doubling time is ~3-5 days. We performed single cell sorting using the BD FACS Aria II into precoated 384-well plates. Approximately 60% of these sorted cells formed colonies (>100 cells) and were frozen ready for further analysis.

These 12 new single cell clonal cultures established from patient SN291 were expanded and DNA and RNA extractions performed for the tumor, 5 quantized cells and PBMC's from patient and family members to complete WGS and whole transcriptomic analysis via RNA-seq. We have been working on cells from patient SN243 and have established similar clonal cultures and also submit these for WGS analysis by Complete Genomics. Once complete, the data will be sent to us and the integrative analysis will be performed and completed at ISB.

Sample preparation for whole genome sequencing (WGS)

Genomic DNAs have been extracted from specimens of both SN291 and SN 243 patients and consenting family members (Figure 2). We have quantified and checked the quality of each of these DNA preps according to protocol provided by Complete Genomics (Mountain View, CA). Ten samples from patient SN291 and family members have been selected for WGS analysis and eleven samples have been selected for WGS analysis for SN243. The samples for SN291 include deeper sequencing of tumor tissue, normal sequence coverage of primary tumor cell culture, 5 subclones, patient PBMCs, and PMBCs from the patient's two children. The samples for SN243 include tumor tissue, primary tumor cell culture, 5 subclones, patient PBMCs, and PMBCs from the patient's two children. The samples have

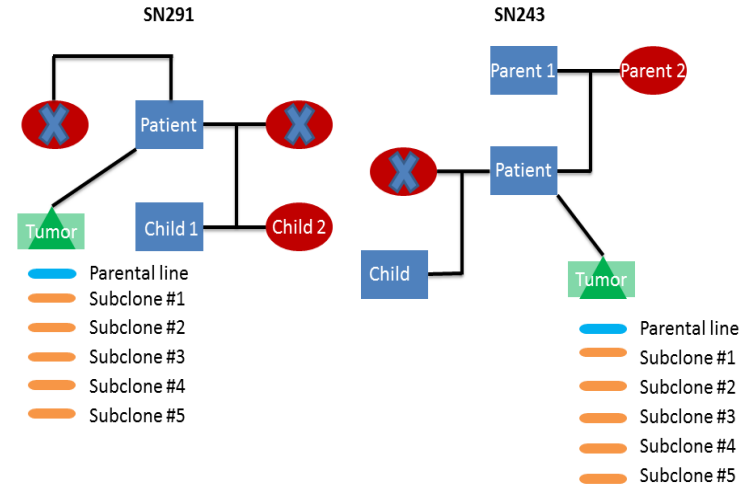


Figure 2. Patient selection and family history of samples submitted to Complete Genomics for WGS analysis.

Sample ID	Customer Sample ID	Customer Subject ID	Tumor Status	QC Status	Details	DNA Vol. Reported by Customer (µl)	DNA Vol. Measured by CGI (µl)	DNA Conc. Reported by Customer (ng/µl)	DNA Conc. Measured by CGI (ng/µl)	Available DNA (µg)	Gender Reported	Count of ChrY SNPs	Total Count of Called SNPs	Gender Match to Reported Gender
GS02741-DNA_A01	SN243 PBMC	PBMC	Non-Tumor	Passed		190	162	78.6	77.8	12.6	Male	9	87	Match
GS02741-DNA_B01	SN243 TISSUE	Tissue	Tumor	Passed		100	88	149.3	167.9	14.8	Male	9	91	Match
GS02741-DNA_C01	SN243 P1	parent 1	Non-Tumor	Passed		70	53	212.8	253.9	13.5	Male	9	91	Match
GS02741-DNA_D01	SN243 P2	parent 2	Non-Tumor	Passed		88	64	170.5	171.7	11.0	Female	0	79	Match
GS02741-DNA_E01	SN243 C1	child1	Non-Tumor	Passed		200	174	61.6	65.6	11.4	Female	0	66	Match
GS02741-DNA_F01	SN243 PARENTAL	parental cell	Tumor	Passed		50	46	200.0	140.8	6.4	Male	9	86	Match
GS02741-DNA_G01	SN243 CLONE 1	clone 1	Tumor	Failed	Quantity Failed	50	43	296.0	56.9	2.4	Male	9	82	Match
GS02741-DNA_H01	SN243 CLONE 2	clone 2	Tumor	Passed		50	44	100.0	87.7	3.8	Male	9	90	Match
GS02741-DNA_A02	SN243 CLONE 4	clone 4	Tumor	Passed		50	44	85.8	92.1	4.0	Male	9	87	Match
GS02741-DNA_B02	SN243 CLONE 6	clone 6	Tumor	Passed		180	169	53.1	45.1	7.6	Male	9	82	Match
GS02741-DNA_C02	SN243 CLONE 7	clone 7	Tumor	Passed		50	43	138.4	200.6	8.6	Male	9	86	Match
GS02741-DNA_D02	SN243 CLONE 12	clone 12	Tumor	Passed		90	78	70.4	54.1	4.2	Male	9	78	Match

Table 2. CGI quality control template for 12 genomic DNA samples from SN243 family.

been submitted to Complete Genomics and are now in the pipeline waiting to be sequenced.

A total of 22 DNA samples have been submitted to Complete Genomics for WGS analysis. All ten DNA samples from patient SN291 and family members passed quality control by CGI, and high quality WGS data have been produced and sent back to ISB for full genome analysis (Figure 3). It proved to be more challenging to extract quality DNAs from several of the clonal cell populations from patient SN243 despite repeated attempts. We finally managed to collect sufficient amount of DNAs from 12 samples in the SN243 cohort. As shown in Table 2, eleven out of twelve samples passed initial CGI quality control, and were moved down their pipeline for sequencing analysis. More recently, we were informed by CGI that DNA samples from two

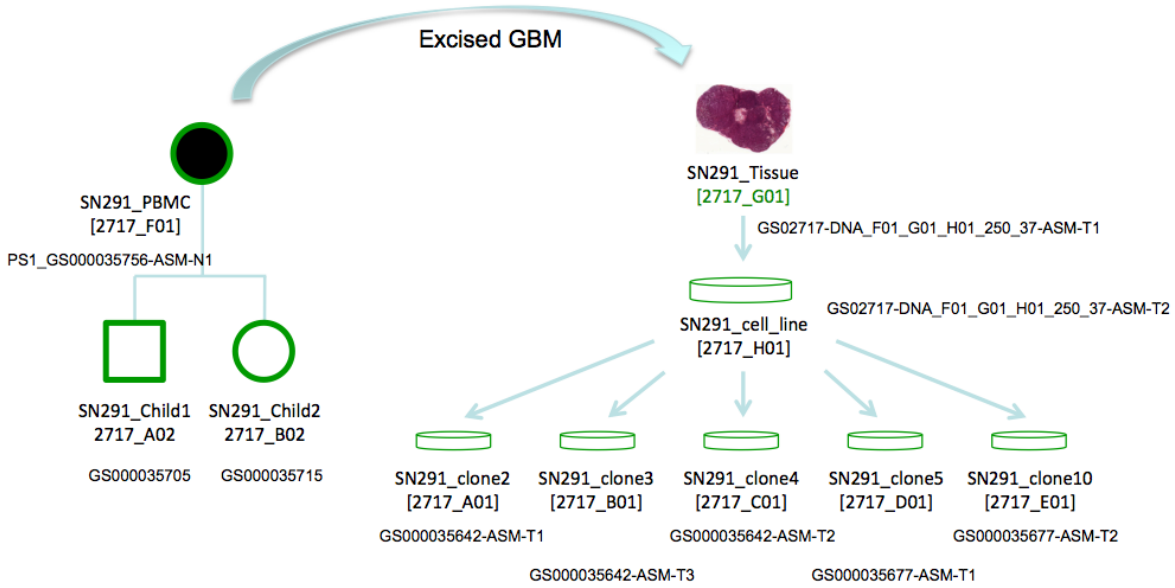


Figure 3. Overview of patient SN291 genome dataset. For each sample, we indicate its descriptive identifier, the vendor sample identifier between square brackets, and the vendor assembly identifier.

clonal populations (4 and 12) failed insertion of one of the adapters, and that CGI has requested more genomic DNAs to be submitted for those two clones. We had not previously encountered these problems before and have been in intensive discussions to rectify this issue as production of quantized cells is a long, laborious and costly step. Regardless, we have started to expand these two clones to generate more DNA for each of the cell subtypes for resubmission to CGI.

Analysis of Whole genome data: Karyotype computed from genome data

We have developed a method for precise identification of aneuploidies at high resolution – from a few kb long to entire chromosomes – based on comparison of the genome coverage signal to a pre-computed "median coverage profile" of many genomes sequenced with the same technology and processed using equivalent pipeline versions. For analyzing the SN291

genomes, we generated a median coverage profile based on 106 normal genomes, all obtained from blood samples, and excluding the currently studied genomes. We then normalized the SN291 genomes to this median profile, and used a hidden Markov model to segment the normalized coverage signal to identify regions of coverage that is lower or higher than expected. Finally, we filtered the resulting segments based on population frequencies, as assessed based on thousands of complete genomes available to us.

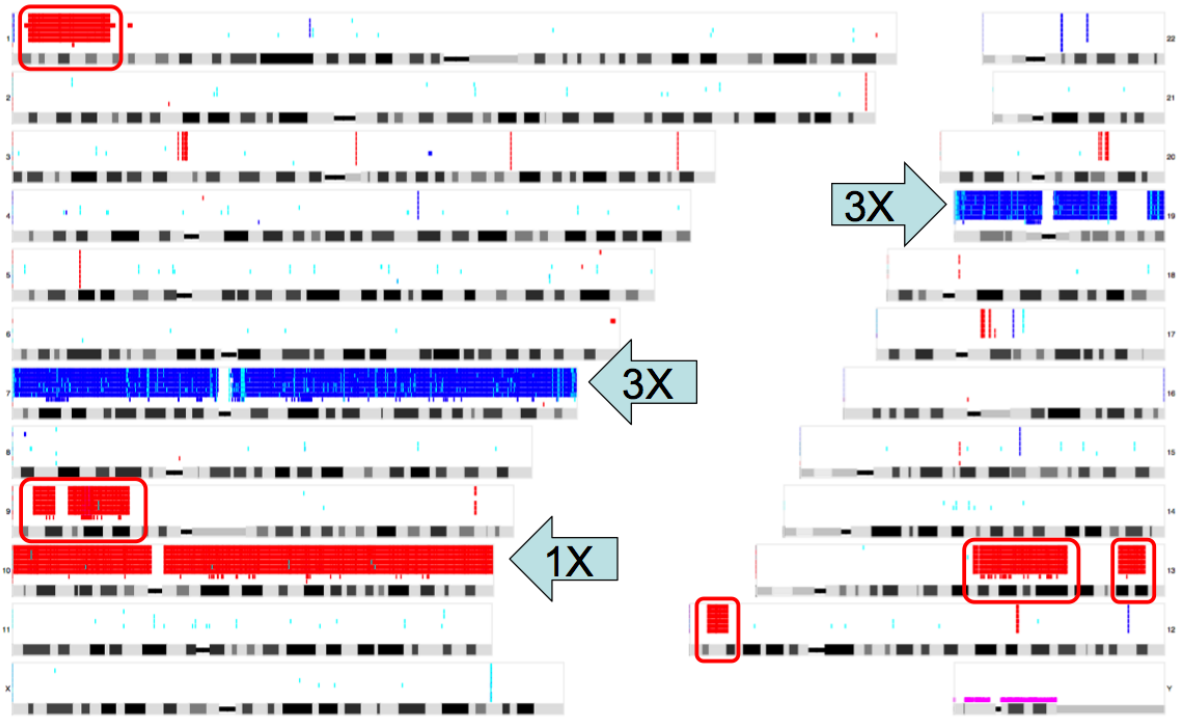


Figure 4. Computed karyotype. For each chromosome, we depict (bottom to top) the computed copy numbers observed for SN291's PBMC genome, cancer tissue, cell line and subclones. Red denotes deletions (haploid), blue represents expansions (triploid, with light blue representing tetraploid or higher).

We applied our coverage analysis algorithm to identify regions of ploidy change in the SN291 genomes. Figure 4 depicts our findings genome-wide. We observed complete loss of one copy of the entire chr10 in the cell line and the subclones. The fragmented signal for the cancer tissue indicates the presence of this aneuploidy in a subpopulation of the mixed tissue. The chromosomal deletion is not observed in the PBMC sample, as expected. Similarly, we observed large-scale but partial losses in chromosomes 1, 9, 12 and 13. Conversely, we observed an extra copy (triploidy) of chr7 and most of chr19.

Relationship between subclones of SN291

Based on the aneuploidy analysis, we could determine that the five subclones are independent of each other as expected (Figure 5). Each of them presents a small number of minor private aneuploidies, none of which is shared by two or more subclones.

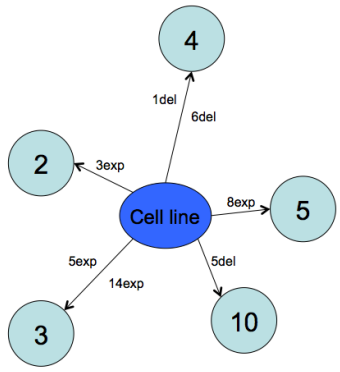
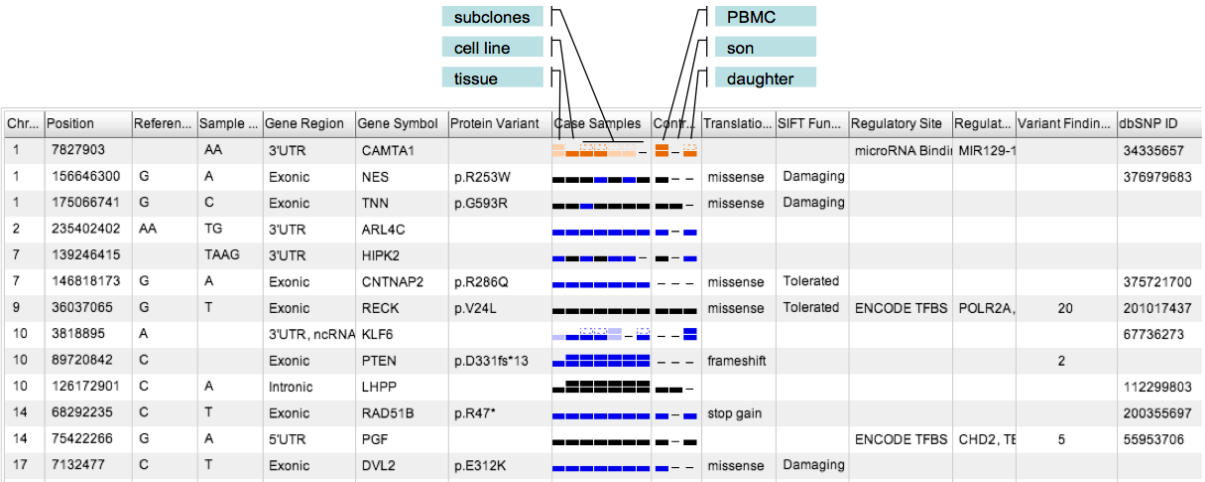


Figure 5. Relationships between the subclones and the cell line. Each subclone presents a very small number of private aneuploidies (deletions or expansions).

Variant analysis

We used "Ingenuity Variant Analysis" to compare the ten genome sequences to identify candidate variants associated with the glioblastoma phenotype, using the tissue, cell line and five subclones as "cases". We required candidate variants to be predicted deleterious, observed in at least three "cases" with quality >= 35, and with population frequency under 1%. We used Ingenuity's knowledgebase to select cancer driver variants directly affecting genes known to be involved in GBM. As shown in Figure 6, we identified a number of interesting gene mutation candidates. Of particular interest is a stop-gain SNV in the RAD51B gene, present in heterozygous form in the genome of the patient (PBMC), the cancer tissue, the cell line and all the subclones. This variant is very rare, with a population frequency of 0.0079% (as computed using our Kaviar genome database) and is confirmed by its presence in the daughter (but absent in the son). A second variant of interest is a novel missense SNV in DVL2, predicted to be deleterious**.**



Chr...	Position	Referen...	Sample ...	Gene Region	Gene Symbol	Protein Variant	Case Samples	Contr...	Translatio...	SIFT Fun...	Regulatory Site	Regulat...	Variant Findin...	dbSNP ID
1	7827903		AA	3'UTR	CAMTA1						microRNA Bindi	MIR129-1		34335657
1	156646300	G	A	Exonic	NES	p.R253W			missense	Damaging				376979683
1	175066741	G	C	Exonic	TNN	p.G593R			missense	Damaging				
2	235402402	AA	TG	3'UTR	ARL4C									
7	139246415		TAAG	3'UTR	HIPK2									
7	146818173	G	A	Exonic	CNTNAP2	p.R286Q			missense	Tolerated				375721700
9	36037065	G	T	Exonic	RECK	p.V24L			missense	Tolerated	ENCODE TFBS	POLR2A,	20	201017437
10	3818895	A		3'UTR, ncRNA	KLF6									67736273
10	89720842	C		Exonic	PTEN	p.D331fs*13			frameshift				2	
10	126172901	C	A	Intronic	LHPP									112299803
14	68292235	C	T	Exonic	RAD51B	p.R47*			stop gain					200355697
14	75422266	G	A	5'UTR	PGF						ENCODE TFBS	CHD2, TE	5	55953706
17	7132477	C	T	Exonic	DVL2	p.E312K			missense	Damaging				

Figure 6. Identification of detrimental variants. Orange and blue denote gain of function and loss of function, respectively.

RNA-Seq Data Analysis

Algorithm selection

To fully extract gene expression information from glioblastoma tumor tissues and clonal populations for biomarker discovery, we have evaluated a number of published computational algorithms for analyzing RNA-seq data. As shown in Figure 7, these algorithms were purposely devised with particular alignment and mapping applications in mind, and each carries a specific strength. Given that our single cell RNA-seq data is of lower coverage, we have decided to use TopHat as our method-of-choice, based on its more general application, robust performance, and widely acceptance in the community.

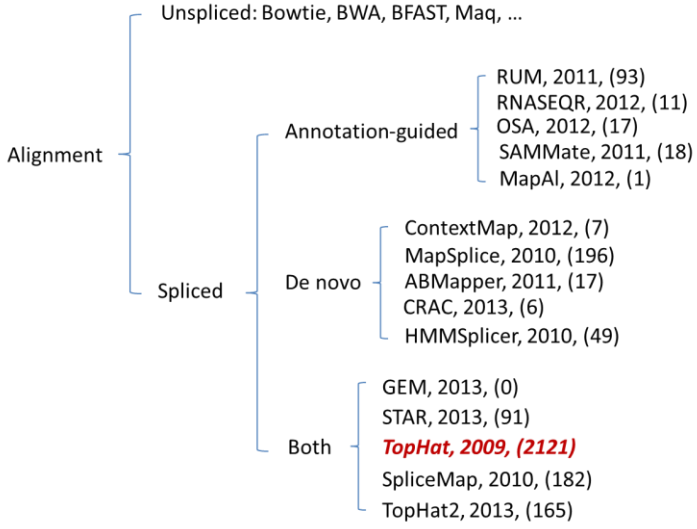


Figure 7. Evaluation of published RNA-seq data analysis algorithms. Numbers in parentheses indicate number of citations for each program.

Single Cell RNA-Seq data for cultured SN291 tumor cells.

We have generated single cell RNA-Seq data from cultured primary tumor cells derived from glioblastoma patients. As shown in Table 3. cDNA libraries from 96 sorted individual cells derived from SN291 patient were prepared, indexed and subjected for next generation sequencing on an Illumina HiSeq platform. Approximately 1-2 million quality reads from 87 cells were generated; ~60% of reads can be mapped concordantly to the human genome. In nine cells, not enough sequencing reads were produced, thus they were excluded from further analysis.

Principle component analysis (PCA) and network mapping.

We conducted PCA analysis on the 87 single cell transcriptomes for SN291. As shown in Figure 8, several distinct cell clusters can be identified. We further mapped enrichment pattern for the CD133+ gene signatures (containing 89 genes) that we published before (PNAS, 2011). Cells bearing CD133+ signature (red) show distinct separation from those cells negative for the signature. One cell (purple) shows a strong enrichment for Wnt signaling pathway genes.

Sample	AlignedPairs	ConcordantPairs	DisconcordantPairs	C1 type	Sample	AlignedPairs	ConcordantPairs	DisconcordantPairs	C1 type
g01	899,235	49.7%	34.0%	1	g49	1,326,584	66.1%	19.7%	1
g02	1,191,278	47.5%	36.3%	1	g50	1,709,060	63.1%	21.8%	1
g03	1,612,721	62.1%	20.9%	1	g51	1,161,220	65.7%	21.2%	1
g04	615	39.5%	43.1%	1	g52	874,909	47.9%	35.6%	1
g05	1,647,110	59.9%	23.6%	1	g53	1,119,424	55.0%	28.8%	1
g06	1,797,059	60.4%	22.1%	1	g54	1,211,442	51.8%	33.1%	1
g07	1,686,501	66.0%	18.6%	1	g55	1,215,567	53.4%	30.6%	1
g08	1,351,292	55.6%	23.4%	2	g56	1,142,071	53.9%	29.1%	1
g09	1,605,567	61.6%	21.5%	1	g57	1,324,289	55.3%	28.1%	1
g10	1,019,528	49.0%	33.4%	2	g58	1,711,576	66.0%	17.7%	1
g11	1,176,268	49.4%	34.2%	1	g59	1,426,244	64.9%	20.5%	1
g12	1,079,971	57.7%	26.6%	1	g60	1,330,170	65.6%	18.6%	1
g13	1,572,060	61.4%	24.6%	1	g61	1,706,541	64.0%	22.5%	1
g14	1,009,804	56.1%	28.1%	1	g62	1,682,185	64.9%	20.4%	1
g15	1,161,573	57.2%	26.6%	2	g63	1,372,796	56.0%	25.8%	1
g16	1,101,282	56.9%	25.7%	1	g64	1,291,539	66.3%	16.3%	1
g17	1,040,458	71.5%	13.6%	1	g65	1,006,918	60.5%	23.2%	1
g18	1,912,866	64.7%	18.3%	2	g66	546,529	50.9%	33.4%	1
g19	1,197,388	57.9%	26.1%	3	g67	1,292,106	63.8%	21.1%	1
g20	1,187,574	55.1%	28.0%	1	g68	211,436	79.7%	8.1%	1
g21	1,611,582	53.7%	26.5%	1	g69	1,306,388	55.0%	28.9%	1
g22	1,419,808	58.4%	24.9%	1	g70	922,079	47.6%	36.3%	1
g23	1,348,420	56.6%	26.6%	1	g71	1,107,861	51.5%	32.4%	1
g24	787,207	59.6%	24.3%	1	g72	257,779	74.8%	10.3%	2
g25	1,460,197	65.5%	19.9%	1	g73	701,982	44.3%	39.4%	1
g26	2,016,625	66.0%	20.3%	1	g74	1,152,883	62.6%	22.9%	1
g27	1,968,362	62.2%	23.0%	1	g75	1,355,419	69.1%	19.5%	1
g28	1,138,651	57.9%	23.0%	1	g76	914,087	63.1%	21.6%	1
g29	1,312,753	57.2%	27.8%	1	g77	1,418,660	70.3%	16.3%	1
g30	1,631,678	63.2%	21.4%	1	g78	1,250,458	58.7%	26.5%	1
g31	1,472,455	65.8%	18.2%	1	g79	1,669,264	65.9%	18.4%	1
g32	1,974	77.0%	15.1%	0	g80	1,036	80.6%	11.1%	0
g33	1,429,980	61.6%	23.6%	1	g81	1,304,525	59.4%	25.0%	1
g34	1,125,710	55.4%	30.7%	1	g82	1,708,550	62.9%	21.7%	1
g35	999,367	52.8%	32.0%	1	g83	1,493,367	68.9%	18.1%	2
g36	1,244,461	56.0%	25.4%	1	g84	508,605	56.4%	26.8%	1
g37	1,619,458	62.1%	22.5%	1	g85	279,937	71.8%	10.0%	1
g38	3,827,684	74.7%	11.5%	1	g86	731,840	47.7%	35.7%	1
g39	2,130,139	67.8%	15.8%	1	g87	980,536	51.1%	33.0%	1
g40	19,340	78.7%	7.4%	0	g88	1,585	82.6%	9.7%	0
g41	1,741,414	68.7%	17.1%	1	g89	1,145,143	47.6%	35.0%	1
g42	1,482,177	68.8%	18.0%	1	g90	1,706,828	65.2%	18.0%	1
g43	1,283,590	59.6%	25.3%	1	g91	1,439,922	65.7%	20.4%	2
g44	1,175,265	55.7%	27.6%	1	g92	1,196,507	58.4%	22.5%	1
g45	1,643,689	63.7%	18.8%	1	g93	1,605,522	66.5%	16.1%	1
g46	2,080,139	65.8%	16.6%	1	g94	1,698,874	71.6%	13.1%	1
g47	1,203,681	52.6%	31.8%	1	g95	1,925,914	62.6%	21.3%	1
g48	1,243,135	56.4%	28.9%	1	g96	1,117	83.6%	10.2%	0

Table 3. RNA-seq analysis of 96 single cells from patient SN291

Refinement of RNA samples from clonal populations for RNA-seq analysis

It will be key to compare the gene expression profiles to the WGS data generated from the same clonal populations. Since April, we have been working to complete the sample preparation for RNA-seq from GBM patients SN243 and SN291. We have extract total RNAs from the parental cell lines plus five to six of their subclones, and have validated the integrity and concentration of the isolated RNAs. We had originally planned to submit the RNA for RNA-seq without generating the sequencing libraries prior to sample submission, however the rising costs that occurred just prior to submission made us halt our plans for that. Instead, we will

generate the libraries ourselves and we are currently waiting to purchase the Illumina sequencing kits to finish the sample prep for the RNA-seq and then we will pool the samples for the sequencing run before sample submission.

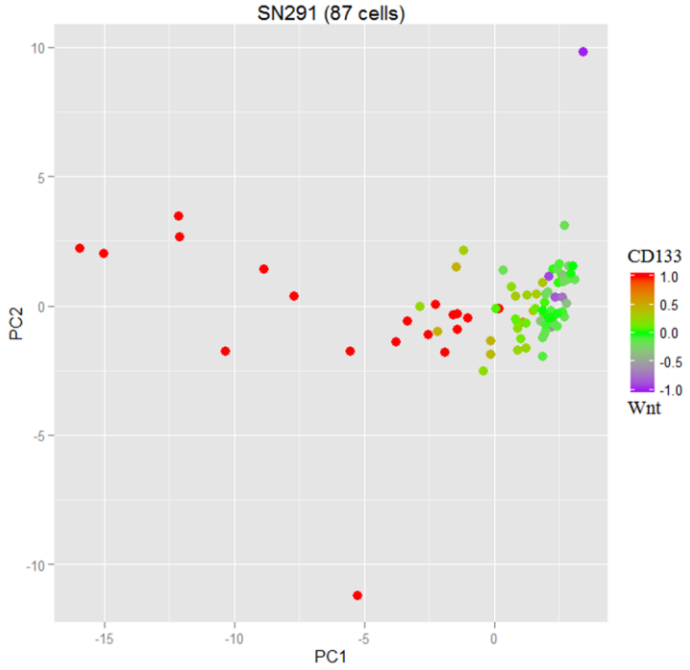


Figure 8. Principle component analysis (PCA) of RNA-seq data generated from individual cells from SN291 tumor culture. Each dot represents a single cell. Color gradient indicates enrichment score for either our published CD133+ gene signature (PNAS, 2011) or Wnt pathway genes.

Aim 3.

We continued our establishment and further growth of the parental cell lines and subclones for proteomic analysis as described above and have developed protocols for the stringent analysis of these samples incorporating genomic information obtained in WGS for the establishment of candidate protein biomarkers. We developed new growth conditions for these cells suitable for proteomic analysis that resulted in no cell morphology changes to allow for the elimination of extraneous protein added from additional cell growth components and protein derived from fetal bovine serum (FBS) added to the culture medium of these cells. Elimination of this protein source allows us to identify proteins directly secreted from these quantized cell populations. We finalized the collection of cell pellets and secreted protein preparations for cell line SN291 and five subclones derived from SN291 (SN291-SC2, SN291-SC3, SN291-SC4, SN291-SC5, SN291-SC10) in this year (Figure 9). Cells were grown in standard media conditions established to provide quantized cells. For the analysis of secreted proteins, cells were grown in FBS free medium for 24h prior to collection in unsupplemented medium. In agreement with the cell lines selected for WGS, the primary culture from patient SN243 was grown for the analysis of the proteome and six subclones were established (SN243-SC1, SN243-SC2, SN243-SC4, SN243-SC6, SN243-SC7, SN243-SC12) to prepare cell pellets and secreted protein fractions. One subclone was grown as alternative as SN243-SC12 was a particularly slow growing clone. Further a media blank was prepared as control. Primary cultures from additional glioblastoma patients, patient SN260 and SN348, were established and classified as potential alternative as addressed

previously. Even though these two patient samples were not subjected to WGS, we included the parental cell lines for comparative proteomic analysis. In summary, we prepared cell pellets and secreted protein sample from four parental cell lines and eleven subclones derived from two of these parental lines (Figure 2 & 9).

We have begun deep proteome analysis of the secretomes by performing off-gel fractionation of the peptide digests of each of the secretome preparations and have combined the analysis of each of the fractions to prepare lists of differentially secreted proteins from each of the cell lines. A table on the progress of the proteomic results is shown (Table 4). The cell pellets will been further processed to study the cell surface proteome using the ISB developed N-glyco capture technology and subsequent analysis by liquid chromatography-mass spectrometry (LC-MS/MS). The N-glycosylated proteins are of interest in the context of biomarker identification strategies and expected to provide insight in differentially expressed signatures of glioblastoma

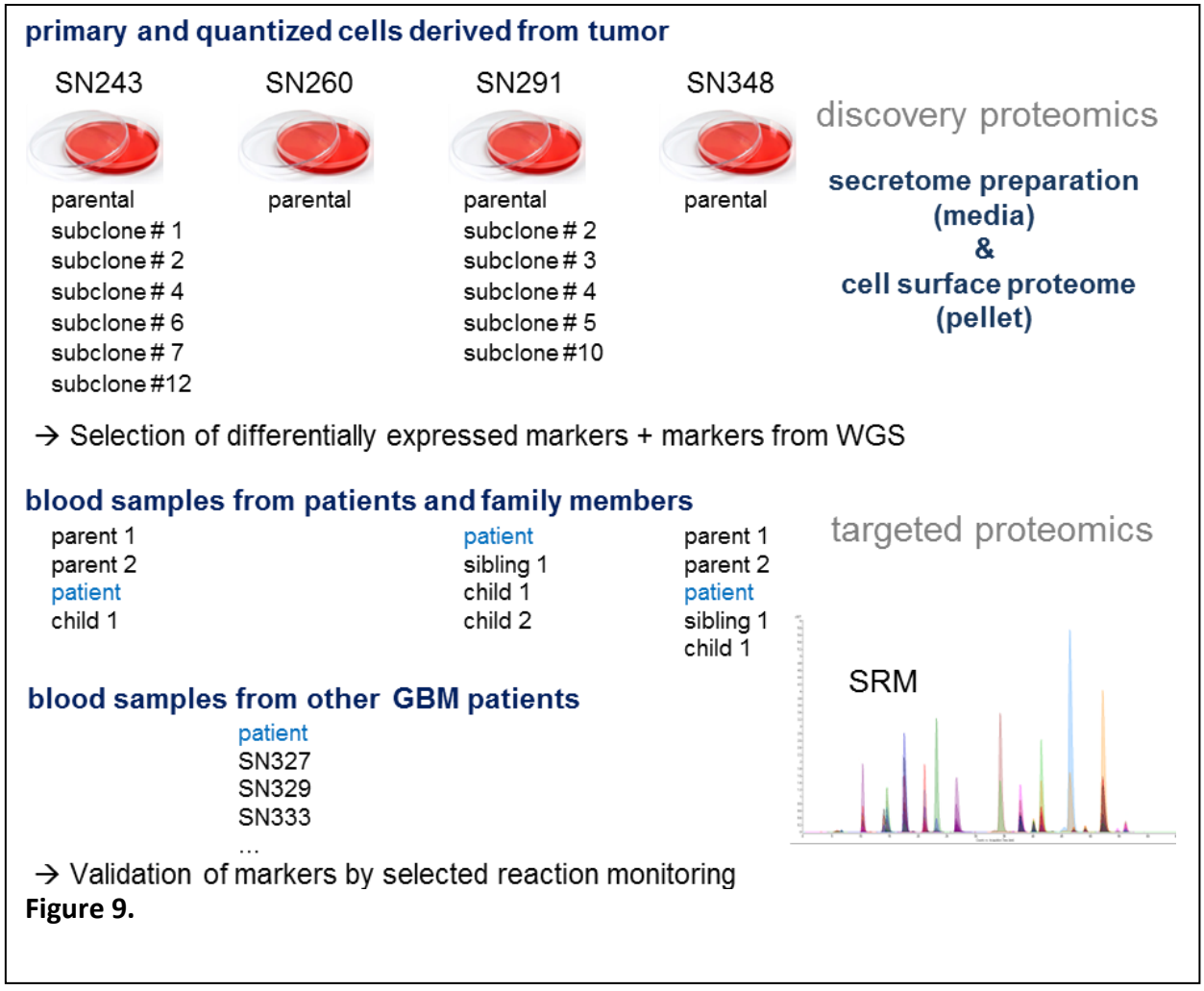


Figure 9.

patients. The secretome profiles of the individual samples are of equal importance to provide candidate markers. Our developed secretome protocol proofed to be successful and the

secretome preparations of the four parental cell lines and eleven subclones yielded enough material for subsequent LC-MS/MS analysis. The 15 samples were subjected to tryptic digestion using a standard protocol including the reduction of disulfide bonds with dithiothreitol and alkylation of the sample with iodoacetamide. All samples were analyzed in duplicate using a high resolution QExactive mass spectrometer (Thermo Fisher Scientific) allowing peptide fragmentation by higher-energy collisional dissociation. Peptides were separated on a reversed phase column using a particular long gradient of 4h and nano-LC conditions to allow highly sensitive in depth analysis of the secretome of each tumor cell sample.

The generated proteomic data are analyzed through sequence database searching using the software tool suite of the Trans-Proteomic-Pipeline (developed at ISB) for the correct assignment of MS spectra to peptides and to infer the proteins from these peptide

Proteomics - Secretome Analysis of Parental Cell Lines and Subclones							
cell line	**type**	**cell culture**	**sample preparation**	**digestion**	**LC-MS/MS analysis**	**data analysis**	**comment**
SN291	parental	✓	✓	✓	✓	in progress	
media blank 1		✓	✓	✓	✓	in progress	
SN291_noS	parental	✓	✓	✓	✓	in progress	supplementals further removed
SN291	subclone 2	✓	✓	✓	✓	in progress	
SN291	subclone 3	✓	✓	✓	✓	in progress	
SN291	subclone 4	✓	✓	✓	✓	in progress	
SN291	subclone 5	✓	✓	✓	✓	in progress	very slow growing
SN291	subclone 10	✓	✓	✓	✓	in progress	
SN243	parental	✓	✓	✓	✓	in progress	
media blank 2		✓	✓	✓	✓	in progress	
SN243_noS	parental	✓	✓	✓	✓		supplementals further removed
media blank 3		✓	✓	✓	✓		
SN243	subclone 1	✓	✓	✓	✓	in progress	
(green = subclones selected for proteomics)	subclone 2	✓	✓	✓	✓	in progress	
	subclone 3	✓					
	subclone 4	✓	✓	✓	✓	in progress	
	subclone 5	✓					
	subclone 6	✓	✓	✓	✓	in progress	
	subclone 7	✓	✓	✓	✓	in progress	
	subclone 8	✓					
	subclone 9	✓					
	subclone 10	✓					
	subclone 12	✓	✓	✓	✓	in progress	
SN260	parental	✓	✓	✓	✓	in progress	
SN348	parental	✓	✓	✓	✓	in progress	

Table 4. Progress on proteomic secretome analysis of parental cell lines and subclones.

identifications. A standard database would allow us to detect known proteins but not the detection of mutational changes from the tumor genome or the tumor derived quantized cells. To include such mutations, we are in the progress of generating an extended cancer genome specific database that considers the results from the whole genome sequence and allows for a correlation of specific mutations arising from these tumor cells on the proteome level. This will be done first for all samples from SN291 (we have full genome sequence information for SN291 now in-house) and subsequently for SN243 once WGS results are received from Complete Genomics.

Biomarker candidates derived from this discovery proteomic analysis will be correlated with the data derived from the transcriptome and whole genome analysis to define the final list of candidates that will be subjected to targeted quantitative proteomic selected-reaction monitoring (SRM) analysis. Blood samples have been collected from patient SN291 and SN243 and of three direct family members from each patient. We also have additional blood plasma samples from patient SN348 and four direct family members as well as other patients (not selected for WGS) to arrive at an extended sample population for targeted SRM analysis of the selected candidates. This effort will enable us to evaluate GBM specific tumor markers in this cohort of patients as well as the larger pool of 100 plasma samples from both GBM patients and normal subjects as defined by our collaborators at SNI.

Database construction for cancer derived mutational proteome analysis.

The standard method for identifying proteins in a mass spectrometry experiment involves the use of a whole proteome database to compare sequence information, which is processed *in silico* and compared to the experimentally observed spectra. The database is meant to represent every possible polypeptide sequence fragment from the subject organism, to afford the best chance of correctly interpreting each experimental spectrum. The quantized cell glioblastoma project has identified numerous mutations from WGS (see Aims-1-4), many of which would encode novel polypeptide sequences that would not be identifiable using traditional proteomics approaches. Since it is clearly infeasible to consider every possible rearrangement of the human genome and resultant modified peptides, we devised a method to encode these variable sequences in a modified whole proteome database in a manner that will enable us to detect the fragmentation spectra from such modified peptides.

Essentially any novel polypeptide sequence resulting from observed genetic rearrangements are appended to the canonical sequence for that particular gene product, with a reasonable amount of flanking sequence as context, to account for missed enzymatic cleavages. Each ‘cassette’ consisting of a modified sequence plus context is separated from each other and the original sequence by an asterisk character. The asterisk is treated as a hard-stop boundary by most search engines, as well as the TPP software used to interpret the results in a statistically

valid manner, so there is no chance of introducing spurious mutations. This allows us to encode virtually all likely sequences in a relatively compact and non-redundant manner, both desirable qualities to keep the search times reasonable and limiting the protein inference problem. We have begun to construct this database compendium from the existing WGS data from SN291 and will further supplement this with EGS data from SN243 once we obtain this.

Aim 5.

Our collaborators at SNI have developed a drug screening protocol for primary GBM cell lines (Hothi et al., 2012). In this study, SN143 and SN186 parental cells lines were tested against a library of 2,000 compounds composed primarily of FDA approved compounds (50%), natural products (30%), and other bioactive components (20%) (MicroSource Spectrum collection, MicroSource Discovery Inc.), to identify inhibitors of GBM stem cell (GSC) proliferation. As shown in Table 5, this drug screening assay was able to identify approximately 100 compounds that were cytotoxic against SN143 and SN186. The compounds identified represent multiple classes of drugs and natural products, including antineoplastics, cardiotonics, antihelminthics, and others. For Aim 5, our goal is to apply this methodology to quantized cells and use this procedure to identify therapeutic agents with the potential to target the stem cell population. In particular, we will compare the drug responsiveness of the parental cultures with the corresponding sub clone populations. We hope this will provide valuable information that can be translated to the clinic and used to design effective treatment strategies for future GBM patients. Given that SN243 and SN291 samples have been used for whole genome sequencing, we will apply a similar drug screening approach to both these primary cell lines and their corresponding subclones. Taking

Class[a]	Total in class	Active agents	
		SN143	SN186
Alcohol antagonist	3	1	1
Antihelminthic	33	7	7
Antiarrhythmic	24	0	1
Antibacterial	227	11	11
Antifungal	55	5	5
Antineoplastic	115	29	28
Antihyperlipidemic	12	3	4
Antihypertensive	63	1	1
Anti-infective	11	2	3
Antipsychotic	22	1	1
Cardiotonic	14	10	10
Diuretic	16	0	0
H1 antihistamine	11	1	0
Immunosuppressant	5	1	0
Psychotropic	9	0	1
Sclerosing agent	2	1	1
Vasodilator	35	0	0
Undetermined activity	444	19	22
Total agents	1101b	92	105

Table 5. Pharmacological classes for inhibitors of GSC proliferation of SN143 and SN186.

into consideration the number of cells required for drug screening and the growth kinetics of the quantized cells, we have decided to pursue a 160-oncology focused library. This library is composed of FDA approved antineoplastics as well as compounds in late phase clinical trials, several of which include glioblastoma-relevant drugs such as PI3K/ mTOR inhibitors, VEGFR inhibitors and met-inhibitors.

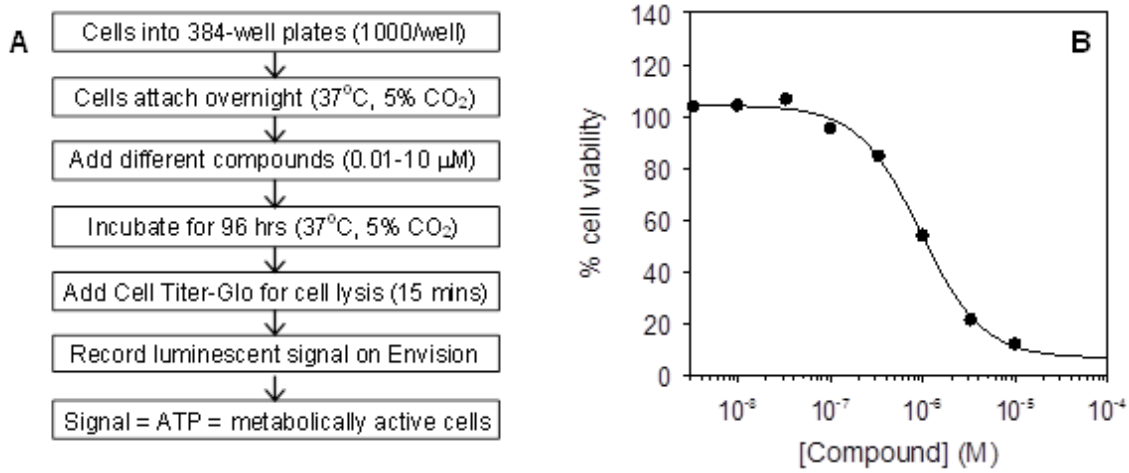


Figure 10. (A) Established protocol for high throughput chemical screens. Viable cells (metabolically active) are determined by ATP quantification, using the CellTiter-Glo luminescent cell viability assay (Promega). **(B)** Example of dose response curve used to determine IC50 value for each drug candidate.

To date, we have optimized drug screening assays with the 160-compound library using the SN186 parental cell line as shown in Figure 10. Drug potency is typically assessed by determining the half maximal inhibitory concentration (IC50) (i.e. the concentration of a drug that is required for 50% inhibition in vitro). With this in mind, we generated 8-point dose response curves to access the IC50 of each drug (Figure 10B). This same technique will be applied to SN243 and SN291 parental cell lines, and their corresponding quantized cell populations in the final year.

Progress:

We have encountered a significant logistical problem with the explosion of whole genome sequencing requests and capacity by our contractor Complete Genomics Inc. We have addressed these issues with them and are progressing towards the completion of this project. With large changes in staff at CGI, it has taken time to reestablish an error free system and we

have been subjected to these delays. Other technical problems we encountered were solved and we do not anticipate problems with the proposed work schedule for the next 12 months. Our efforts in whole genome sequencing have advanced since or request for a no-cost extension of the proposal and bode well for the upcoming work to be completed. This program will be b completed as stated and has benefited from further significant technical developments we have made and will provide significant results over the program in the next 12 months.

KEY RESEARCH ACCOMPLISHMENTS FOR 2013-2014

- We have established 6 GBM primary cell lines suitable for extensive molecular analysis with three of these cell lines (SN291, SN243 and SN348) for which there is full family consent and both blood and tumor was collected from these patients, blood from family members collected and stored as PBMC's and plasma that are suitable for both cellular analysis, genomic analysis and proteomic analysis of these patients.
- Quantized cell populations have been established from Samples SN291 and SN243.
- Obtained WGS of patient SN291, tumor, primary cell line, 5 separate quantized cells and SN291 family members. Prepared DNA and submitted samples for SN243 WGS.
- Identified key mutations in the cancer genome of the SN291 GBM sample
- Transcriptomic profiling of single cells from quantized populations of SN291 GBM tumors. Expression patterns of selected genes among these single tumor cells that will digitally stratify tumor cells into distinct populations have been established. Prepared mRNA samples for SN243 RNA-seq.
- Developed cell culture conditions for secretome analysis of quantized cells and protein extraction conditions to maximize the amount of protein for high-mass accuracy quantitative mass spectrometry.
- Performed deep proteome secretome analysis of individual GBM primary cells and their quantized subclones totaling 18 different cells to date
- Constructed targeted lists of initial proteome identifications from GBM tumor samples and constructed cancer proteome specific database strategies to identify protein mutations predicted by WGS.
- Established a drug screening protocol for GBM primary cells and quantized cell populations with preliminary results from patient samples SN143 and SN186 reported in the previous year for technical development work.

REPORTABLE OUTCOMES

We have reported our work emanating from the efforts described here in this report in several publications describing some of our technical developments applied to GBM cell analysis.

Hothi P, Martins TJ, Chen L, Deleyrolle L, Yoon JG, Reynolds B, Foltz G. High-throughput chemical screens identify disulfiram as an inhibitor of human glioblastoma stem cells. Oncotarget. 2012 Oct;3(10):1124-36. PMID: 23165409.

Sangar V, Funk CC, Kusebauch U, Campbell DS, Moritz RL, Price ND. Quantitative proteomic analysis reveals effects of EGFR on invasion-promoting proteins secreted by glioblastoma cells.

Mol Cell Proteomics. 2014 Jul 5. pii: mcp.M114.040428. PMID: 24997998.

CONCLUSION

Description of work to be performed during the next reporting period.

Over the next 12 months starting in July 2014 to June 2015 we will concentrate on the following efforts as described in our statement of work:

We will complete our whole genome sequence collection for patient and family SN243. This will complete our data collection for whole genome sequencing and provide a second extensive dataset to supplement the efforts we have made significant inroads on. We have established a data analysis pipeline wit patient SN291 and their family members and will complete this analysis.

We will complete our full transcript analysis of single quantized cells from both SN291 and SN243 to establish expressed gene and mutations identified in the whole genome sequence analysis of these sample types.

We will complete our proteomics analysis of the full sample set of expanded GBM patients and of the family members from two of these undergoing full genomic analysis. We will perform tumor proteome specific analysis of GBM primary and quantized cell lines to provide quantitative differences between these as well as detect and quantify expressed protein mutations identified from the multi-omic approach.

We will compile a candidate biomarker list of select targets derived from WGS, single cell RNA-seq and cancer proteome data of both whole cells as well as quantize cell secretomes to build targeted assays for candidate biomarker evaluation. We will construct a panel of SRM assays and analysis conditions to deploy across GBM patient and normal subject plasma samples to conduct a biomarker evaluation for early detection of GBM tumors in both a unblended and blinded analysis.

We will complete our targeted drug screening profile of GMB patient primary and quantized cells to reveal inhibitors of GBM stem cell proliferation and compile a list of potential drugs that can be utilized in Phase-II clinical trials.

REFERENCES

N/A

APPENDICES

N/A